IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

VOICE ACTIVATED/VOICE RESPONSIVE ITEM LOCATOR

Inventor: Jerome R. Mahoney

Attorney Docket No. IVC-103A

Kenneth P. Glynn, Esq.
Attorney for Applicant
Reg. No. 26,893
Glynn and Associates
Suite 201 (Plaza One)
One Route 12 West
Flemington, NJ 08822-1731
tele (908) 788-0077
fax (908) 788-3999

007081700KPG-hd

VOICE ACTIVATED/VOICE RESPONSIVE ITEM LOCATOR

Attorney Docket No: IVC-103A

5   BACKGROUND OF THE INVENTION

1.   Field of the Invention

The present invention relates to voice

activated/voice responsive item locators, i.e.

item directories which direct a user such as a

10   consumer or shopper, to a specific location to

view, treat, retrieve, order, purchase or

otherwise use the information obtained in the

system. Typically, the present invention could be

used at retail stores to locate items to be

15   purchased. Alternatively, it could be used at a

production facility or distribution facility

having a large number of parts, to locate

specific parts for an employee. In other

embodiments, it could be used in non-commercial

entities, such as a public library to locate a

particular book. The locator of the present

invention relies upon a specific software module

5   to accomplish voice recognition and response, and

includes manager programming for customization,

updates and modifications.

2. Information Disclosure Statement

The following prior art patents represent

10   various inventions relating to machines involving

speech recognition for voice-based operation and

thus illustrate known voice recognition

applications:

U.S. Patent No. 5,111,501 to Masanobu

15   Shimanuki describes a telephone terminal device

equipped with a transmitter microphone, a

receiver, a speech recognition unit that receives

and recognizes speech signals from the

transmitter microphone and a circuit to reduce

the level of signals send from a telephone

5   network to the receiver when the speech

recognition unit receives speech signals from the

transmitter microphone. Further, this device is

preferably equipped with a speech reproduction

unit that reproduces the speech information

10   stored in a memory, in response to the

information of recognition result from the speech

recognition unit, and a circuit that prevents

transmission of signals from the telephone

network to the receiver when the regenerated

15   speech information is sent to the receiver.

Furthermore, it is desirable for this device to

be provided with a circuit that prevents

generation of ringing tones when an incoming call

arrives.

U.S. Patent No. 5,136,634 to David C. Rae et

5   al. describes voice operated facsimile machine

network which includes a method and apparatus for

transmitting specifically requested graphic

and/or textual data from an unattended database

storage location to a requestor's facsimile

10   machine over a telephone line which includes a

host computer such as a PC modified with a

facsimile transmission board and a voice

generation board. The host computer receives

incoming phone calls and prompts the caller using

15   the voice board to select data files by using the

DTMF keys of a standard telephone handset. The

PC can be left unattended and can run

automatically in the facsimile transmission mode.

Callers can immediately access needed textual and

image data with the use of just a standard

5   telephone and facsimile machine. Multiple

workstation nodes can be configured in a network

setup to handle a high volume of calls in real

time and to allow multiple data services to

operate simultaneously.

10   U.S. Patent No. 5,165,095 to Mark A.

Borcherding describes a method for dialing a

telephone, using voice recognition to initiate

the dialing and to determine the correct

telephone number. The dialing is initiated with

15   a spoken dial command that is recognized by using

speaker independent templates that are stored

locally with respect to the caller's telephone.

The correct telephone number is recognized by

using speaker dependent template that are

downloaded from a central database or by using

5   speaker independent templates stored locally.

     U.S. Patent No. 5,168,548 to Steven Kaufman

et al. describes a reporting system which is

disclosed herein, a speech recognizer which is

used to select selections of text from a report

10   form stored in a computer and to insert

recognized terms in the text thereby to generate

a report text under voice control. A command

interpreter, also responsive to spoken words,

initiates creation of the report text and its

15   subsequent storing, printing and transmission.

The command processor is responsive to respective

spoken commands to select a destination telephone number and to cause the report text to be sent to apparatus for converting report text to image data and for modulating an audio band signal with

5   the image data for facsimile transmission over telephone lines.

U.S. Patent No. 5,222,121 to Keiko Shimada describes a voice recognition dialing unit of a telephone mounted on a vehicle or similar mobile

10   body and which allows a call to be originated with ease. When the user of the telephone enters a voice command on voice inputting section, the dialing unit originates a call automatically and thereby connects the other party to the telephone

15   line. In a call origination procedure, the operations for call origination and the

verifications are performed between the user and

the unit in an interactive sequence. In a

preferred embodiment, the unit has a particular

call origination procedure in which, when the

other party recognized by the unit is wrong as

determined by the user by verification, lower

place candidates for the other party are called

up in response to a particular voice command. In

an alternative embodiment, the unit indicates the

other party by voicing a name for verification

purpose. The alternative embodiment selects and

stores only the name of the other party in

response to an entered voice signal and, in the

event of response for verification, combines the

name having been stored and response information

stored beforehand to produce composite response

voice.

U.S. Patent No. 5,231,670 to Richard S.
Goldhor et al. describes a system and method for
generating text from a voice input that divides

5   the processing of each speech event into a

dictation event and a text event. Each dictation

event handles the processing of data relating to

the input into the system, and each text event

deals with the generation of text from the

10   inputted voice signals. In order to easily

distinguish the dictation events from each other

and text events from each other the system and

method creates a data structure for storing

certain information relating to each individual

15   event. Such data structures enable the system

and method to process both simple spoken words as

well as spoken commands and to provide the

necessary text generation in response to the

spoken words or to execute an appropriate

function in response to a command. Speech

5   recognition includes the ability to distinguish

between dictation text and commands.

U.S. Patent No. 5,239,586 to Kuniyoshi Marui

describes a voice recognition system which

comprises a handset and a hands-free microphone

10   for generating an input audio signal, a high-pass

filter for eliminating low frequency components

from the signal from the handset or hands-free

microphone, a signal lever controller for

adjusting the level of the high-pass signal in

15   response to the user of either the handset or

hands-free microphone, a storer for storing the

speech data and a controller for controlling the

storer so that a user's utterance is stored or

the user's utterance is recognized by comparing

the utterance to speech data already stored. The

5   handset hook switch provides an on-hook control

signal to reduce amplifier gain during hands-free

microphone operation.

U.S. Patent No. 5,301,227 to Shoichi Kamei

et al. describes an automatic dial telephone that

10   is useable in a motor vehicle, when a voice input

is provided during a period in which input of the

names of called parties is awaited, a voice

pattern of the name of the called party is

compared with reference patterns of called

15   parties stored in reference patterns storing

device, to determine the degree of the similarity

therebetween. The names of the called parties

are output to a user in the order of decreasing

degree of similarity. Each time the name of a

called party is output, a command word for

5   confirmation is a waited from a user for a

predetermined time period. When a voice

confirmation command is input and is recognized

during this waiting period, a telephone number

corresponding to the name of the called party is

10   supplied to a channel. Consequently, the command

word for confirmation may be input only if the

name of the called party outputted is one desired

by the user. Sensors continually monitor the

driving condition of the motor vehicle in which

15   the telephone is installed. When the operation

of the steering wheel or brakes of the motor

vehicle exceeds a predetermined threshold or the speed of the motor vehicle is excessive, the sensors generate safety signals that inhibit the operation of the telephone.

U.S. Patent No. 5,335,276 to E. Earle Thompson et al. describes a communication system which is provided with multiple purpose personal communication devices. Each communication device includes a touch-sensitive visual display to communicate text and graphic information to and from the user and for operating the communication device. Voice activation and voice control capabilities are included within communication devices to perform the same functions as the touch-sensitive visual display. The communication device includes a built-in modem,

audio input and output, telephone jacks and

wireless communication. A plurality of

application modules are used with personal

communication devices to perform a wide variety

5   of communication functions such as information

retrievable, on-line data base services,

electronic and voice mail. Communication devices

and application modules cooperate to allow

integrating multiple functions such as real time

10   communication, information storage and

processing, specialized information services, and

remote control of other equipment into an

intuitively user friendly apparatus. The system

includes both desktop and hand-held communication

15   devices with the same full range of communication

capabilities provided in each type of

communication device.

U.S. Patent No. 5,349,636 to Roberto Irribarren describes a communication system for verbal telephonic communication which has a voice message system for storing and retrieving voice messages integrated with a computer database accessing system for storing and retrieving text messages from a separate computer system and for converting the text messages into voice. The systems are integrated via a network which coordinates the functions of each individual system. Additionally, the input/output ports of the voice message system and the computer database accessing system are connected in a parallel fashion to at least one telephone line. In this configuration a user may access both

voice messages and database information,

including text or electronic mail messages, with

a single telephone call. Optionally, facsimile

messages can be stored, retrieved and manipulated

5   with a single telephone call.

U.S. Patent No. 5,406,618 to Stephen B.

Knuth et al. describes a telephone answering

device that is activated by a proximity sensor

when a user crosses its field of detection and

10   whose operation is controlled by simple voice

commands. The device incorporates speaker-

independent voice recognition circuitry to

respond to spoken commands of the user that are

elicited by a system generated voice request

15   menu. The telephone answering device performs

all the basic functions of a telephone answering

machine in response to these simple commands and there is no need for the user to manually operate the telephone answering device.

U.S. Patent No. 5,602,963 to W. Michael

5 Bissonnette et al. describes a small, portable, hand-held electronic personal organizer which performs voice recognition on words spoken by a user to input data into the organizer and records voice messages from the user. The spoken words

10 and the voice messages are input via a microphone. The voice messages are compressed before being converted into digital signals for storage. The stored digital voice messages are reconverted into analog signals and then expanded

15 for reproduction using a speaker. The organizer is capable of a number of different functions,

including voice training, memo record, reminder,

manual reminder, timer setting, message review,

waiting message, calendar, phone group select,

number retrieval, add phone number, security and

5   "no" logic. During such various functions, data

is principally entered by voice and occasionally

through use of a limited keypad, and voice

recordings are made and played back as

appropriate. A visual display provides feedback

10   to the user. During the various function, the

user can edit various different data within the

organizer by eliminating or correcting such data

or entering new data.

U.S. Patent No. 5,621,658 to Brion K.

15   Jackson describes an action contained within an

electronic mail object which is communicated from

a data processing system to another data

processing system via an audio device. The

action is executable on a data processing system.

At the sending data processing system, the action

5   is converted to a predetermined audio pattern.

The electronic mail object may contain text in

addition to an action. The text is also

converted to an audio pattern. The audio

patterns are then communicated to the audio

10   device over telephone lines or other

communication medium. At the receiving end, the

audio device records the object. A user can

provide the recorded object to a data processing

system, which then executes the action and

15   converts the text audio patterns back to text.

In addition, the action can be converted to text

and displayed on the data processing system.

U.S. Patent No. 5,631,745 to John J. Wong et al. describes a telephone terminal adapted for business or home use that includes the ability to

5   receive and send facsimiles, a voice answering function and a computer modem. Various input and output devices may be used for the facsimile function. A voice annotated facsimile may be sent and received. At the same time the

10   facsimile is viewed on a video monitor or ordinary television set, an accompanying voice message is heard through the sound system of the monitor or television set. The terminal has an architecture including a central processor and an

15   internal bus structure to which several types of memory, various input-output devices and an

interface with the telephone line are connected,

among others. Audio Random Access Memory (ARAM)

is used for storing both facsimile data and voice

data.

5   U.S. Patent No. 5,671,328 to Gregory P.

Fitzpatrick et al. describes a method and data

processing system which are disclosed for

automatically creating voice processing template

entries. In one embodiment, the invention

10   automatically assembles a plurality of commands

received by the data processing system, at least

one of said commands having a voice recognition

criteria component associated therewith, counts

the occurrences of the plurality of commands,

15   assembles voice recognition criteria components

associated with the plurality of commands, and,

as a result of the occurrence count exceeding a

predefined minimum, constructs a voice

recognition template entry by associating the

assembled voice recognition criteria components

5   with the assembled plurality of commands.

        U.S. Patent No. 5,850,627 to Joel M. Gould

et al. describes a word recognition system which

can: respond to the input of a character string

from a user by limiting the words it will

10   recognize to words having a related, but not

necessarily the same, string; score signals

generated after a user has been prompted to

generate a given word against words other than

the prompted word to determine if the signal

15   should be used to train the prompted word; vary

the number of signals a user is prompted to

generate to train a given word as a function of

how well the training signals score against each

other or prior models for the prompted word;

create a new acoustic model of a phrase by

5 concatenating prior acoustic models of the words

in the phrase; obtain information from another

program running on the same computer, such as its

commands or the context of text being entered

into it, and use that information to vary which

10 words it can recognize; determine which program

unit, such as an application program or dialog

box, currently has input focus on its computer

and create a vocabulary state associated with

that program unit into which vocabulary words

15 which will be made active when that program group

has the focus can be put; detect the available

computational resources and alter the

instructions it executes in response; test if its

ability to respond to voice input has been shut

off without user confirmation, and, if so, turn

5   that ability back on and prompt the user to

confirm if that ability is to be turned off;

store both a first and a second set of models for

individual vocabulary words and enable a user to

selectively cause the recognizer to disregard the

10   second set of models for a selected word; and/or

score a signal representing a given word against

models for that word from different word model

sets to select which model should be used for

future recognition.

15

Notwithstanding the prior art, the present

invention is neither taught nor rendered obvious thereby.

SUMMARY OF THE INVENTION

A voice activated/voice responsive item locator system is disclosed to enable a user to speak into the system and have the system respond with location information for an item requested by the user. For example, shopper at a home supply store may pick up a locator phone or just speak into a wall mounted or otherwise situated microphone and say "Locate Outdoor Paint" or "Find Hammers" or simply state what is sought without the use of a verb, e.g. "Caulking". The system may reply either with voice or visual (words on a screen, or map), or both voice and visual, e.g. "Aisle 3, Shelf 4". In some

instances the system will reply, for example,

with a "Repeat", or "Restate in different words"

or "Please talk to information desk" or other

default instructions.

5   The locator system may be a stand alone

device, but in most embodiments would be part of

an internal connected system. It could be an

intranet or secured internet system, but would in

many cases be a storewide system with a plurality

10   of user locations (units, phones, or microphones,

with feedback at each location). The system will

include an embedded voice-driven interface for

speech control of: (1) operational instructions;

(2) core system locator function operations, that

15   is, recognition of specific requests and

responses thereto; and, (3) optional and default

functions. In preferred embodiments, the present

invention device is both operated by speech

(speech or voice activated) and speech responsive

(voice answers and instructions to the user from

5   the system). Thus, the present invention device

relies upon automatic speech recognition (ASR),

either in place of or in addition to manual

locator systems, e.g. book, list, map and

computer directories. In some embodiments, user

10   feedback features are included wherein both audio

and visual feedback is given to a user in

response to recognizable voice signals, while in

other possible embodiments, the user may

designate audio or visual.

15   BRIEF DESCRIPTION OF THE DRAWINGS

The present invention should be more fully

understood when the specification herein is taken

in conjunction with the drawings appended hereto

wherein:

Figure 1 shows a general schematic diagram

5   showing software and functional features of a

present invention item locator system;

Figure 2 shows a schematic diagram

illustrating the physical functions of a present

invention voice recognition item locator device;

10   and,

Figure 3 shows a schematic diagram of a

present invention device illustrating details of

a voice recognition submodule.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

15   The present invention is a voice

activated/voice responsive item locator and

system. By "item" is meant a place or thing that a user desires to locate. Thus, a item could be a particular brand of canned string beans, a type of outdoor stain, a booth at a convention, a

5 particular part in inventory for sale, assemblage or distribution, a particular automobile in a production facility lot or in a large parking garage, or a room, afunctional group or a person in an office building or the like. The response

10 may be in the form of a word or sentence presented visually or audibly and it may designate an aisle, a shelf, a bin number, a rom number, a row and slot or space, etc.

The voice recognition system digitizes words

15 spoken via a receiver (microphone) handset, headset, or built-in microphone for conversion

from analog to digital utilizing a continuous

speech recognition digital signal processor

(DSP). The main support structure may be a

conventional type housing for phones and other

5   communications devices, may be of a different

shape or configuration or may be built into a

device such as a wall or desk unit, with or

without monitor. They could be portable or

permanently affixed and could be powered by any

10   means available, e.g. AC or DC current. In the

portable mode, the system would be wireless for

the user and would, in that respect operate like

a cell phone, two way radio, "walkie talkie" or

other short distance wireless device, but would

15   have a processor at a central or fixed location

having the same features as described above,

i.e., the DSP with programming capabilities, etc.

The DSP is connected to a programmable microprocessor and either by customized input or a standard program, the system enables the user

5   to quickly enter voice-activated fields, e.g., such as "Where is...", "Find...", etc.

Verification of voice recognition accuracy (prior to execution) is optional and may be accomplished via synthesized voice playback and/or a screen

10   confirmation which requires a "YES" or "NO" to execute or open for revision. In some preferred embodiments, a screen, e.g., LCD, enables visual feedback during input phase, with support for deletion, insertion, correction, etc.

15   Cancellation of the entire command or programming instructions may be possible at any time (prior

to execution), via keystroke or voice command.

The essential features of the present

invention involve the creation of a voice based

guide or locator to offer enhanced convenience

5   and speed to users for location of one or more

items.

Figure 1 shows a general schematic diagram

of a present invention system showing general

software features and functional features. Thus,

10   the present invention device includes a central

processor 1 which may be an external or internal

component, i.e., within a single unit or at a

separate location from audio receivers and

transmitters , e.g., microphones/speakers for

15   user inputs and feedback to users.

The system may be preprogrammed with the

user being required to follow concise

instructions for activation and operation, or may

be programmable to alter, add or enhance ease or

methods of use, e.g. through a limited access

5   code, for manager inputs 3 of user instructions.

In any event, manager inputs 3 shall include

functional selections and inputs of items and

their locations, with provision for subsequent

access for modifications. This programming may

10   include direct keyboard, voice, etc., and, as

mentioned, may include security capabilities for

preventing unauthorized use, e.g. voice

identification (user recognition) or user

security code system, as well as other options

15   which may be included therein, such as a "help"

detailed manager instruction section.

Once the system has been programmed for
use, the user operation unit(s) 5 provide
functional access, which may be passive, i.e.,
the user speaks, picks up a phone, presses a

5   button, or otherwise takes some action to
activate the system; or it may be active, i.e., a
proximity sensor, a periodicity timer, or other
internal mechanism may automatically activate the
system and could trigger an audio or visual

10   query, such as "May I help you locate a product?"

Once the system has been activated and a
user has stated the necessary words of input to
activate the device, recognition/non-recognition
response 7 results from processing the user

15   inputs to central processor 1 , and audio and/or
video response unit(s) 9 provide feedback 11 to

the user, either by answering the inquiry,

conditionally defaulting, e.g., asking for a

repeat or a restate the question, or fully

defaulting, e.g. directing the user to a courtesy

5   desk or check out counter for help.

Figure 2 shows a schematic diagram

illustrating a present invention voice

activated/voice responsive item locator system,

showing the physical arrangement and function of

10   components. Thus, symbol 17 indicates an

optional user prompter proximity sensor and

symbol 21 is a microphone or equivalent component

for voice input. The voice input is sent to

audio controller 19 and to automatic speech

15   recognition unit 23 and is converted from analog

to digital signals. CPU/Memory 25 compares the

digital signals to the set up or dictionary of

digital words or phrases in memory. Once a match

is found, the system processor 27 and data

storage 31 operate to respond with an answer or a

5   default instruction or a query by providing

digital text to text-to-speech generator 29,

which provides audio feedback to a user via audio

controller 19 and speaker 33. Feedback to a user

may also be provided on visual screen 37 via

10   display controller 35. Keyboard 39 is used for

manager set up and modifications.

Figure 3 shows the details of one preferred

embodiment of the submodule used in the present

invention device. The voice recognition

15   component converts an acoustic signal into a

sequence of labels. The system takes the raw

acoustic data, and processes it through the

recognizer. The recognizer then matches it

against a set of models using a decoder that

generates a recognition token. This token

5   represents what the user said as either a single

word or utterance. The recognizer itself does

not interpret the meaning of the recognized

output, that is the function of the interpreter

(described later). The recognizer uses Hidden

10   Markov Models (HMMs) to provide for a continuous

speech recognition engine. HMMs do not process

the acoustic signal directly but instead split

the signal into a sequence of discrete

observations. These observations are derived

15   from a digital representation of the signal that

had been converted from the analog signal

generated by the microphone.   During recognition,

the likelihood of each model (or sequence of

models) matching the incoming signal is

calculated.   The recognizer simply selects the

5   most likely model to decode the signal.   As this

is done continuously, the recognizer can process

speech as opposed to isolated words, allowing the

user to talk more naturally.

Each acoustic model represents a short

10   sound.   The interpreter combines these sounds

into words using a dictionary.   This dictionary

specifies the pronunciation of each word in terms

of the acoustic models.   After identifying the

most likely word, the interpreter then joins sets

15   of models together (using a Viterbi decoder) in a

series of pre-defined connections such that paths

can be established to provide for a degree of

"natural language" recognition; in other words,

the user can say "Find hammers", "Where are

hammers" or "hammers" and they are all understood

5   to mean the same thing. Moreover, these sets of

models and dictionaries are interchangeable,

allowing the same voice recognition component to

be used in a variety of applications.

As the voice recognition component is

10   running continuously, there needs to be a way to

distinguish background conversations that might

accidentally trigger an unwanted action by the

device. For example, two people standing by a

voice-activated device might be discussing

15   locations of different goods in a supermarket and

be misinterpreted or undesireably responded to.

To avoid this problem, the recognition unit

requires a command word to trigger before

beginning further recognition. The trigger word

is a user-definable setting.

5   Thus, in Figure 3, initialization 51

initiates monitoring 53 for a trigger word from a

user. When a word is received, it is analyzed to

determine whether or not a trigger word 55 has

been received. If not, signal 57 returns the

10   status to monitoring 53 for a new word. This

loop continues until a trigger word is recognized

and an inactivity timer 59 is started. The

monitor 61 proceeds with the monitoring for the

next word and waits for timer pop 65. When an

15   event 63 is received, timer pop 65 returns to the

monitor 53 to continue the monitoring process and

the voice data is sent to interpretation 67. If

it is understood 69, an action 75 if process and

feedback function 77 is performed. Additionally,

signal 79 prompts user 71. Likewise, if the

5   interpretation is not understood 69, user 71 is

prompted and via signal 73, timer 59 begins

again. These cyclings operate on a continual

basis while the system is initiated. Voice

activation may also be used to shut down the

10   system.

Obviously, numerous modifications and

variations of the present invention are possible

in light of the above teachings. It is therefore

understood that within the scope of the appended

15   claims, the invention may be practiced otherwise

than as specifically described herein.